



De Novo Transcriptome Sequencing and Analysis of the Cereal Cyst Nematode, *Heterodera avenae*

Mukesh Kumar¹, Nagavara Prasad Gantasala¹, Tanmoy Roychowdhury², Prasoon Kumar Thakur¹, Prakash Banakar¹, Rohit N. Shukla⁴, Michael G. K. Jones³, Uma Rao^{1*}

1 Division of Nematology, Indian Agricultural Research Institute, New Delhi, India, **2** School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, **3** School of Veterinary and Life Sciences, Murdoch University, Perth, Australia, **4** Bionivid Technology [P] Ltd, Bangalore, India

Abstract

The cereal cyst nematode (CCN, *Heterodera avenae*) is a major pest of wheat (*Triticum spp*) that reduces crop yields in many countries. Cyst nematodes are obligate sedentary endoparasites that reproduce by amphimixis. Here, we report the first transcriptome analysis of two stages of *H. avenae*. After sequencing extracted RNA from pre parasitic infective juvenile and adult stages of the life cycle, 131 million Illumina high quality paired end reads were obtained which generated 27,765 contigs with N50 of 1,028 base pairs, of which 10,452 were annotated. Comparative analyses were undertaken to evaluate *H. avenae* sequences with those of other plant, animal and free living nematodes to identify differences in expressed genes. There were 4,431 transcripts common to *H. avenae* and the free living nematode *Caenorhabditis elegans*, and 9,462 in common with more closely related potato cyst nematode, *Globodera pallida*. Annotation of *H. avenae* carbohydrate active enzymes (CAZy) revealed fewer glycoside hydrolases (GHs) but more glycosyl transferases (GTs) and carbohydrate esterases (CEs) when compared to *M. incognita*. 1,280 transcripts were found to have secretory signature, presence of signal peptide and absence of transmembrane. In a comparison of genes expressed in the pre-parasitic juvenile and feeding female stages, expression levels of 30 genes with high RPKM (reads per base per kilo million) value, were analysed by qRT-PCR which confirmed the observed differences in their levels of expression levels. In addition, we have also developed a user-friendly resource, *Heterodera* transcriptome database (HATdb) for public access of the data generated in this study. The new data provided on the transcriptome of *H. avenae* adds to the genetic resources available to study plant parasitic nematodes and provides an opportunity to seek new effectors that are specifically involved in the *H. avenae*-cereal host interaction.

Citation: Kumar M, Gantasala NP, Roychowdhury T, Thakur PK, Banakar P, et al. (2014) De Novo Transcriptome Sequencing and Analysis of the Cereal Cyst Nematode, *Heterodera avenae*. PLoS ONE 9(5): e96311. doi:10.1371/journal.pone.0096311

Editor: Philippe Castagnone-Sereno, INRA, France

Received: November 19, 2013; **Accepted:** April 7, 2014; **Published:** May 6, 2014

Copyright: © 2014 Kumar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors would like to thank the Department of Biotechnology, Government of India for financial support under the Indo-Australia Strategic Research fund (PR12678). M.G.K. Jones thanks the Australia-India Strategic Research Fund (BF030027) for financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: R. Shukla is an employee of Bionivid Technology [P] Ltd. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: umarao@iari.res.in

Introduction

The extent of crop losses caused by plant parasitic nematodes is substantial and contributes to significant reductions in crop yields resulting in an annual yield losses of about US\$157 billion [1]. The cereal cyst nematode (CCN), *Heterodera avenae* (Wollenweber, 1924), is one of the three economically important cyst nematode species that attack wheat and barley crops in many cereal growing regions of the world [2]. Although agronomic management practices are often deployed to manage plant nematodes, control with chemical nematicides is also used for high value crops. Because chemical nematicides are toxic and persistent, they are a human health risk and most are being phased out [3]. There is therefore an urgent need to find new gene targets which can be used to develop novel and environmentally friendly nematode control methods [1,4].

Cyst nematodes are obligate sedentary endoparasites that reproduce mainly by amphimixis. After infection of host plants, they develop a close interaction with and feed from a group of interconnected syncytial cells of their host plants. The life cycle starts with an egg present inside an encysted female. The first stage larva (J1) develops within the egg and second stage larva (J2)

hatches in response to low soil temperatures (5–15°C) [5]. J2s may invade main or lateral roots in the zone of elongation and migrate intracellularly towards the vascular cylinder [6]. Each J2 selects a single cell that becomes the initial feeding cell (IFC). In a susceptible host, the cells next to the IFC expand and become interconnected after local dissolution of cell walls at pit fields, resulting in the formation of a multinucleate syncytium with characteristics of transfer cells [7]. J2s feed from associated syncytia and undergo three molts, during which they develop either into a mobile male or become a sedentary endoparasitic female that continues to feed until reproduction is completed. The female then dies and forms a cyst which protects the eggs inside it. As a survival strategy not all J2s hatch simultaneously in the same season, but hatching can occur over several years with some unhatched eggs retained within the cyst [8].

Nematodes are amongst the earliest recognised parasites of plants and animals. However, plant parasites have been studied in less detail than many animal parasites and free living nematodes. Most studies have focused on the free living model nematodes, *Caenorhabditis elegans* and *C. briggsae* [9] and comparative studies with such nematodes have been useful in annotating genes and

genomes of plant parasites [10]. To date, genetic resources of plant parasitic nematodes have been based mainly on EST sequences [11] and from annotated genomic sequences of the root-knot nematodes, *Meloidogyne incognita* and *M. hapla* [1,12]. Comparatively less sequence information is available for *Heterodera* spp than for root knot nematodes, exceptions being *H. schachtii* and *H. glycines* [13,14]. In addition, the potato cyst nematode *Globodera pallida* genome has been published [15]. Recently, transcriptomes of two migratory endoparasitic root lesion nematodes, *Pratylenchus coffeae* and *P. thornei* have also been published [16,17].

Next generation sequencing (NGS) has greatly increased our ability to analyse transcriptomes and genomes of many organisms in a cost-effective manner [18]. In this study, using Illumina short reads, we have sequenced, assembled and annotated transcriptomes of two life stages of *H. avenae*, for which very little genomic data was available previously. Data is presented on gene transcripts expressed in two developmental stages of *H. avenae*, and transcript levels have been compared between these developmental stages. The data also provide information on potential targets for control of *H. avenae*. Overall, the *H. avenae* transcriptome provides significant new information on expressed genes, which adds to overall understanding of its parasitic abilities, and which could be used to develop novel management strategies.

Methods

Nematode collection and multiplication on wheat plants

Cysts of *H. avenae* were collected from infested wheat roots growing in the plots of the Indian Agricultural Research Institute, New Delhi, India. Genetic homogeneity of the population was determined by PCR-RFLP and sequencing of ITS (Internal Transcribed Spacer) regions of rDNA (ribosomal DNA). A genetically homogenous population was then multiplied on wheat plants grown in pots kept in a greenhouse [19]. Newly formed cysts were collected manually from the wheat roots using a dissecting microscope and kept for hatching. Hatched J2s were collected and later used for RNA extraction. Likewise, hatched J2s were also used to infect wheat roots and feeding females (FF) that developed and protruded from roots 6–8 weeks after infection were collected and frozen (−80°C) for RNA extraction.

RNA extraction, cDNA synthesis, library preparation and sequencing

Total RNA was extracted from the frozen J2s and FFs using TRIzol (Invitrogen Life Technologies) according to the manufacturer's instructions. Extracted RNA was assessed for quality and quantity using an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA with an RNA integrity number (RIN) of 8.0 was used for mRNA purification (Figure S1). mRNA was purified from 1 µg of intact total RNA using oligodT beads (TruSeq RNA Sample Preparation Kit, Illumina). The purified mRNA was fragmented at elevated temperature (90°C) in the presence of divalent cations and reverse transcribed with Superscript II Reverse Transcriptase (Invitrogen Life Technologies) by priming with random hexamers. Second strand cDNA was synthesized in the presence of DNA polymerase I and RNaseH. The cDNA was cleaned using Agencourt Ampure XP SPRI beads (Beckman Coulter). Illumina adapters were ligated to the cDNA molecules after end repair and addition of an 'A' base followed by SPRI clean-up. The resultant cDNA library was amplified using PCR for enrichment of adapter ligated fragments, quantified using a Nanodrop spectrophotometer (Thermo Scientific) and validated for quality with a Bioanalyzer (Agilent Technologies). It was then sequenced using the Illumina GAIIx platform at Genotypic Technology Private Limited,

Bangalore, India. The sequence data generated has been deposited in the Array Express database for public access (Array Express accession: E-MTAB-2221, ENA accession: ERP004648).

De novo transcriptome assembly and analysis

High quality filtered paired and orphan sequence reads (Phred Score >20) obtained from J2s and FFs were merged and used as an input for assembly of the transcriptome [20]. The Velvet-Oases pipeline was used for *de novo* transcriptome assembly (Velvet assembly is based on de Bruijn graph [21]). Assemblies with different k-mer sizes were used to assess the optimal assembly range. The Oases module was used to generate unique transcripts from the merged assembly. The assembled transcripts with sequence length longer than 100 bp were then considered for transcript annotation and quantification. Assembly was validated by aligning raw reads back to transcripts [20]. Reads were aligned using Bowtie [22] with default parameters.

Annotation and quantification of the transcriptome

Annotation for all the unique transcripts (>200 bp) was done using BLAST [23], homology search against Uniprot [24], the National Center for Biotechnology Information (NCBI)-NR Protein database [25] and NEMABASE4 (<http://www.nematodes.org/nembase4/>). In addition, BLASTX was performed to identify homologues in other databases including Refseq, Swissprot [26], European Molecular Biology Laboratory [27] (EMBL), DNA Databank of Japan [28] (DDBJ), Protein Information Resource [29] (PIR) and Protein Data Bank [30] (RCSB). Nematode orthologues were identified from other completely sequenced genomes by the reciprocal blast method [23]. To study gene orthologues across nematode species, we used the predicted protein sets from seven genomes/transcriptomes available in the public domain (Wormbase, NCBI and Sanger) viz., *C. elegans*, *M. hapla*, *M. incognita*, *Pristionchus pacificus*, *G. pallida*, *P. thornei* and *Ascaris suum*. BLASTP hits with e-value scores ≤1e-5 and query coverage above 60% were considered as annotated homologous proteins and Python script was used for filtering reciprocal best hits. The *Heterodera avenae* transcriptome database (HATDB) is a freely available resource. The database was developed using platform independent and open-source software's MySQL server, PHP, HTML and CSS.

KEGG homologues were identified using the KEGG Automated Annotation Server (KAAS) with default parameters. The resulting BLAST hits were processed by BLAST2GO software [31] to retrieve associated GO terms describing biological processes, molecular functions and cellular components. BLAST2GO uses specific gene identifiers and accession numbers to generate GO annotations as well as corresponding enzyme commission numbers (EC) for sequences with an e-value ≤1e-5. Homologues of the *C. elegans* RNAi pathway and phenotype genes were also identified in the *H. avenae* transcriptome by performing TBLASTX with e-value ≤1e-5. The CAZymes database was searched to find enzymes involved in carbohydrate metabolism by the CAZymes Analysis Toolkit (CAT) [32] with an e-value threshold of 0.01, a bit score threshold of 55 and rule support level of 40. For further confirmation, BLAST and Enzyme Commission results were incorporated. *H. avenae* transcripts were studied for the presence of parasitic effectors identified in other *Heterodera* spp and reported in the published literature. Peptides potentially secreted were identified using SignalP [33] and those with trans-membrane motifs were removed using TMHMM [34]. SecretomeP 1.0 was used to identify non-classical secretion with default parameters [35]. Repeat elements were identified using Repeat Masker version 3.3.0 [36]. Repeat Masker was installed with RMBlast

version 2.2.23+ and Tandem Repeats Finder 4.04 [37]. The transcripts were used as queries against Repbase version 20120418. Default parameters were used with the specification of species as “Nematoda”. Short Sequence Repeats (SSRs) with default parameters were identified with the help of Perl script program MISA (MicroSatellite; <http://pgrc.ipk-gatersleben.de/misa>). Stringent detection criteria were implemented to obtain perfect SSR motifs. Threshold search criteria in MISA script were considered to be appropriate if repeats of mono-nucleotides were present 12 times, di-nucleotides repeated more than 8 times and tri-, tetra-, penta- and hexa-nucleotide repeated 5 times. Expression levels of all the transcripts in the individual libraries (J2 and FF) were assessed by mapping high quality (HQ) filtered reads using BOWTIE2 [38]. Mapped reads were further normalized using RPKM method.

Putative biological significance analysis and Biological Analysis Network (BAN) modeling

Based on the results of transcript quantification, GO Elite tool [39] was used to identify statistically significant regulated GO categories and KEGG pathways with a stringent cut off level of False Discovery Rate (FDR) < 0.05 for regulated transcripts. Transcripts with statistically significant homologies to GO terms and KEGG pathway genes identified by GO Elite tool were used as input for Biological Analysis Network (BAN). Cytoscape v8.1 was used to cluster the genes and processes using edge weighted force directed (Bio-layout). A network analyser plugin was used to identify enriched biological categories and regulated genes. Enriched nodes and edges were network modeled using a hierarchical layout algorithm.

Quantitative Real-Time PCR analysis

Total RNA was extracted from J2s and the FFs of *H. avenae* using a MN NucleoSpin RNA extraction kit (Macherey-Nagel). The quality and concentration of each RNA sample was determined using RNA 6000 Nanokit and a Bioanalyzer (Agilent 2100). Only those samples having RIN values more than 7.0 were used for analysis. cDNA was synthesized from 300 ng of each of the RNA samples using a cDNA synthesis Kit (Superscript VILO, Invitrogen). Quantitative real-time PCR (qRT-PCR) was performed using SYBR Green I as reporter dye in a Realplex² thermal cycler (Eppendorf). A master mix for each PCR reaction was prepared with SYBR Green I, ROX passive reference, stabilizers and PCR Core Reagents (Eurogentec). 1.5 ng of cDNA and 750 nM of each specific primer was added (the primers used in qRT-PCR are listed in Table S1). The following amplification program was used: 95°C 5 min, 40 cycles at 95°C 15 seconds followed by 60°C for 1 min. The *18S rRNA* gene was used as an internal control. The average ct mean values of two biological replicates were taken for calculating the fold change ($2^{-\Delta\Delta CT}$) [40]. Fold change values of highly expressed genes in female were calculated by comparing with J2s and vice versa. There were two biological and three technical replicates for each sample.

Results

Transcriptome sequencing and assembly

mRNA sequencing using the Illumina GAIIx platform yielded a total of 144 million reads of 100 base read lengths from the J2s and FF transcriptomes. Velvet assembly was done with various k-mer ranges and optimal assembly was attained at k-mer 83. Oases tool was used to identify non-overlapping isoforms/splice variants with an insert length of 150 bp and standard deviation of 50 bp. The final assembly resulted in 27,765 transcripts with an average length

of 680 bp and N50 of 1,028 bp (Table 1) [(Raw data statistics) (Table 2)]. Duplicates were then removed by merging transcript assemblies from k-mer-79 to 83. The N50 value was improved after running Oases (1,028): 24,888 contigs (86%) were longer than 200 bp, with 22% of the total contigs between 200–300 bp, 43% between 300 and 900 bp, and 21% longer than 900 bp (Figure S2).

Characterisation of *H. avenae* transcripts

H. avenae transcripts were annotated by comparing the sequences obtained against non-redundant protein sequences available in GenBank and Refseq, Swissprot, EMBL, DDBJ, PIR and RCSB using BLASTX. From the analysis, 10,452 (37.64%) transcripts out of 27,765 were annotated based on the BLASTX score using an e value cutoff of 1e-5 and with minimum query coverage of 50% (Table S2). *H. avenae* transcripts showed significant hits primarily to animal parasitic nematodes (3,387) followed by free living nematodes (2,081) (Figure 1). Some genes known to be important in plant parasitic nematodes such as FLPs (FMRamide-like peptides), MSPs (Major Sperm Proteins), *Expansin*, NLPs (Neuropeptide-like proteins) and *cysteine proteases* were identified amongst the transcripts (Table S2).

The *H. avenae* transcripts were compared with completely sequenced genomes of nematodes using the reciprocal blast method to assign putative orthologues. *H. avenae* sequences shared 6,163 orthologues with *M. incognita*, 6,903 with *M. hapla* and 1,335 with *P. thornei*. It also shared 4,144 orthologues with the free living nematode, *C. elegans*, and 8,315 sequences with the animal parasitic nematode, *Ascaris suum* (Figure 2). The species with the most shared sequences was *G. pallida* with 9,462 common transcripts.

Transcripts present in other nematodes in the form of expressed sequence tags (ESTs) were also identified in *H. avenae* using NEMABASE4 that houses EST data of 62 nematode species [10]. To identify the extent of this similarity, *H. avenae* transcripts were compared with NEMABASE4 annotated EST data sets of all the nematodes using BLASTN (e-value ≤ 1e-5) (Table S3a, b) [11]. The analysis revealed a total of 4,836 significant hits of which 96.42% of the transcripts matched homologues of Heteroderidae according to best BLASTN hits without many hits with other nematodes. This may be due to nucleotide level variation among different group of nematodes [41]. BLAST searches with EST (Table S4) data could annotate 1,839 transcripts which did not show homologues in NR database (Table S3b).

The majority of *C. elegans* genes have been surveyed for knockout phenotypes using RNAi, where matching mRNA is degraded by the introduction of sequence-specific double-stranded RNA (dsRNA) [42]. A comparative analysis was done between *H. avenae* contigs and *C. elegans* genes to identify homologues of genes with RNAi phenotypes in *C. elegans*: 4,147 contigs of *H. avenae* were found to be homologous to 2,543 genes of *C. elegans* with at least one reported RNAi phenotype (Table S5). Primary sequence similarity based searches were then undertaken to identify putative homologous RNAi pathway genes of *C. elegans* in *H. avenae*, and 30 RNAi pathway gene homologues were found in the *H. avenae* transcriptome (Table 3).

Heterodera avenae transcriptome database (HATdb)

The data in the HATDB can be easily accessed. HATDB is available on the URL <http://insilico.iari.res.in>. Users can query the database by providing various options for search of nucleotide and protein like GO (Biological component, Molecular activity and Molecular process) Pfam id. Transcript Identifier. The sequence-based search has been employed using Viro-BLAST

Table 1. Details of raw data and quality control used for assembly of the *H. avenae* transcriptome.

Parameters	Juvenile	Feeding Female
Total Number of Reads	92228826	74250262
HQ Paired End Reads	65264470	66168842
Orphan Reads	10210925	3224372
Low Quality Reads	16753431	4857048

doi:10.1371/journal.pone.0096311.t001

[23,43] search which provides the user friendly option of different BLAST algorithms to search and download nucleotide or protein sequence (s) against *Heterodera avenae* transcriptome sequence data reported in this study. In addition, the links have also been provided to download the whole assembled transcripts and predicted protein.

Putative functional classification using Gene Ontology and KEGG pathway analysis

GO terms were assigned to 10,751 transcripts for further functional characterisation (Figure 3). In this analysis, the predominant transcripts identified as genes involved in catalytic activity constituted 32.7% of the transcripts, of which 10.2% were involved in hydrolytic activity and 0.8% in lyase activity, and 26.6% were identified with binding activity, of which 0.3% were carbohydrate binding. Amongst those categorized under biological processes, transcripts for genes involved in metabolic processes (31.4%) and cellular processes (30.5%) were well represented. Results also indicated that 403 transcripts were homologous to genes involved in cellular communications (Figure 3) [44]. Neuropeptides that are responsible for neuro-muscular activity comprised 0.3% of the transcripts [45]. In addition, 15 transcripts

appeared to be involved in cell cycle regulation (Figure 3) [46]. Metabolic pathways were assigned to *H. avenae* contigs using the Kyoto Encyclopedia of Genes and Genomes database for determining their biochemical functions [47]. Many of the transcripts matched genes in KEGG pathways; metabolic pathways (479), biosynthesis of secondary metabolites (144), oxidative phosphorylation (66), biosynthesis of phenylpropanoids (1) and ubiquitin mediated proteolysis (60). There were 20 KEGG Orthology (KO) mapped to the lipid biosynthesis pathway (Table S6a, b).

Transcriptome quantitation and enrichment of significant biological categories

In silico quantitation of transcripts was done by mapping the reads from individual libraries (J2 and FF) using BOWTIE2 [38], which provided an estimate of transcript abundance. Regulated genes were subjected to GO term enrichment analysis using GOElite V1.2.5 (http://www.genmapp.org/go_elite/). Analysis revealed 64 biological categories under GO terms and pathways significantly enriched at the stringent cut off level of False Discovery Rate (FDR) <0.05 (Table 4). To identify the major functional themes, enriched GO terms were organised into a

Table 2. Assembly statistics of *H. avenae* transcriptome generated by Velvet and Oases.

Assembly statistics of <i>H. avenae</i> transcriptome generated by Velvet				
Kmer length	71	75	79	83
Assembled reads	83346228	80457236	79060568	78113600
Total reads	144868609	144868609	144868609	144868609
Percentage assembled	57.5323	55.5381	54.574	53.9203
Total No. of Contigs	49587	42782	36482	30201
Total assembly length (bp)	22863966	21458075	19925639	18289226
Max contig length (bp)	10420	8935	9073	8997
Min contig length (bp)	141	149	157	165
Average sequence length (bp)	461.088	501.568	546.177	605.583
N50 (bp)	596	666	748	859
Assembly statistics of <i>H. avenae</i> transcriptome generated by Oases				
Parameters	Readings			
Transcripts (Unique transcripts)	27765			
Total Assembly length (Mb)	18.93			
Min transcript length (bp)	101			
Max transcript length (bp)	9075			
Average sequence length (bp)	682			
N50 (bp)	1028			

doi:10.1371/journal.pone.0096311.t002

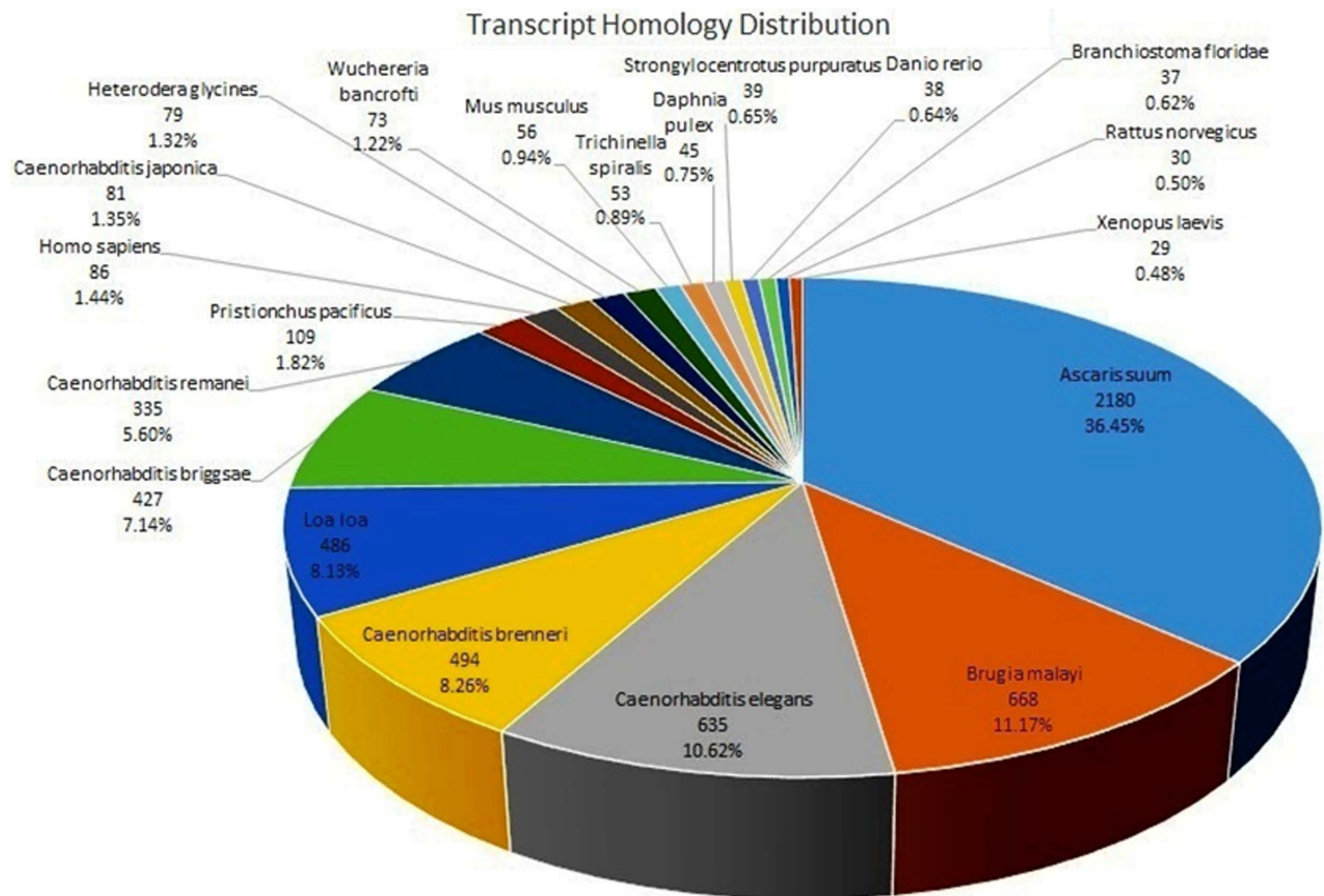


Figure 1. Distribution of the top 20 species with most homologues to *H. avenae*. The distribution was calculated using best BLASTX hit. doi:10.1371/journal.pone.0096311.g001

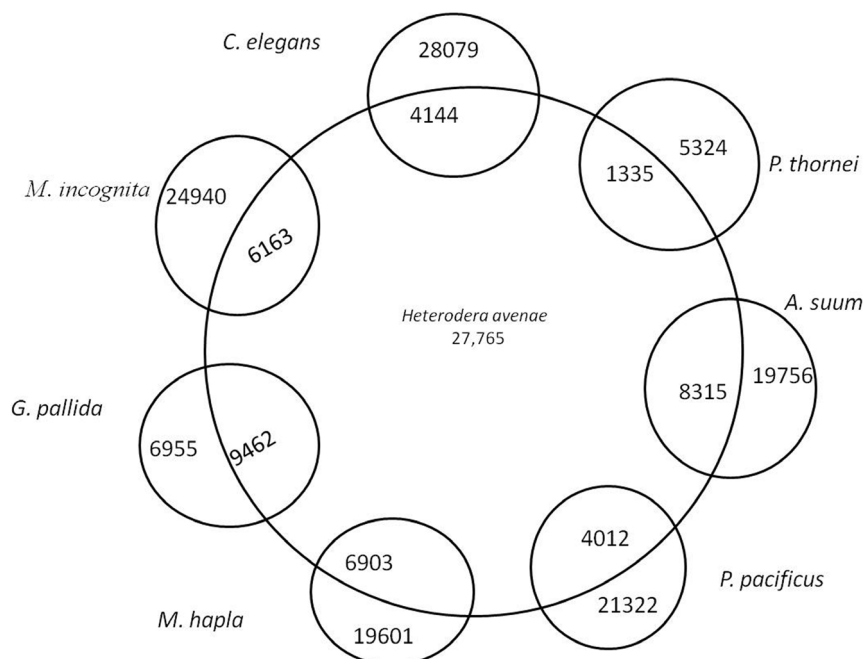


Figure 2. *H. avenae* orthologues present in selected completely sequenced genomes of free living (*C. elegans*, *Pristionchus pacificus*), animal parasitic (*A. suum*) and plant parasitic (*M. incognita*, *M. hapla*, *G. pallida*, and *Pratylenchus thornei*) nematodes. The numbers represent the common sequences. doi:10.1371/journal.pone.0096311.g002

Table 3. Comparative analysis of putative RNAi pathway genes in *H. avenae*, *C. elegans* and *M. incognita*.

<i>C. elegans</i>	<i>H. avenae</i>	<i>M. incognita</i>
Dicer complex		
dcr-1	+	+
drh-1	+	+
drh-3	+	+
xpo-1	+	+
xpo-2	+	+
drsh-1	+	+
rde-4	-	+
rde-5	-	+
RISC complex		
alg-1	+	+
alg-2	-	+
alg-4	+	+
ain-1	+	+
vig-1	-	+
tsn-1	+	+
RO6C7.1	-	+
CO4F12.1	-	+
F58G1.1	-	+
T22H9.3	-	+
csr-1	+	+
sago-1	+	-
ppw-1	-	-
ppw-2	+	-
nrde-3	+	-
RdRp amplification complex		
ego-1	+	+
rrf-1	+	+
rrf-3		
Systemic RNAi (spreading)		
rsd-2	-	-
rsd-3	+	+
rsd-6	-	-
sid-1	-	-
sid-2	-	-
Required for RNAi		
zfp-1	+	+
smg-2	+	+
smg-6	-	-
mes-8	-	+
mes-6	-	-
mes-2	+	+
mut-2	+	-
mut-7	+	-
gfl-1	+	-
cid-1	+	+
ekl-4	-	+
ekl-1	+	+
rha-1	+	+
gfl-1	+	-

Table 3. Cont.

<i>C. elegans</i>	<i>H. avenae</i>	<i>M. incognita</i>
RNAi inhibitors		
eri-1	+	+
eri-7	+	-
xrn-2	+	+

doi:10.1371/journal.pone.0096311.t003

functional network with 34 nodes and 14 edges using enrichment map plugin in Cytoscape v 8.1 (Figure 4). Force directed layout visualisation of the enrichment map showing FDR-based sub clustering revealed four functional clusters including; a) carbohydrate binding b) proteolysis c) motor activity d) hydrolase activity, hydrolyzing O-glycosyl compounds (Figure 4). Visualisation of sub-clusters using hierarchical layout suggested that hydrolase activity was up-regulated in J2s whereas binding activity was up-regulated in FFs.

Validation of putative differentially expressed genes by qRT-PCR analysis

qRT-PCR was performed to identify some differentially expressed transcripts by taking cue from transcript quantitation (Table S2). Thirty genes were selected for further study based on statistical significance, annotation and presence in at least one developmental stage (J2 or FF) of *H. avenae*. Amongst them, 15 genes were expressed highly in the FF stage viz., (*Spectraplakins*, *Ran-BPM*, *CAEBREN_25536*, *CAEBREN_07082*, *Receptor family ligand binding region containing protein*, *Chitinase-like protein*, *Superoxide dismutase*, *Oxidoreductase dhs-27*, *'Paired box' domain containing protein*, *Chitin binding*

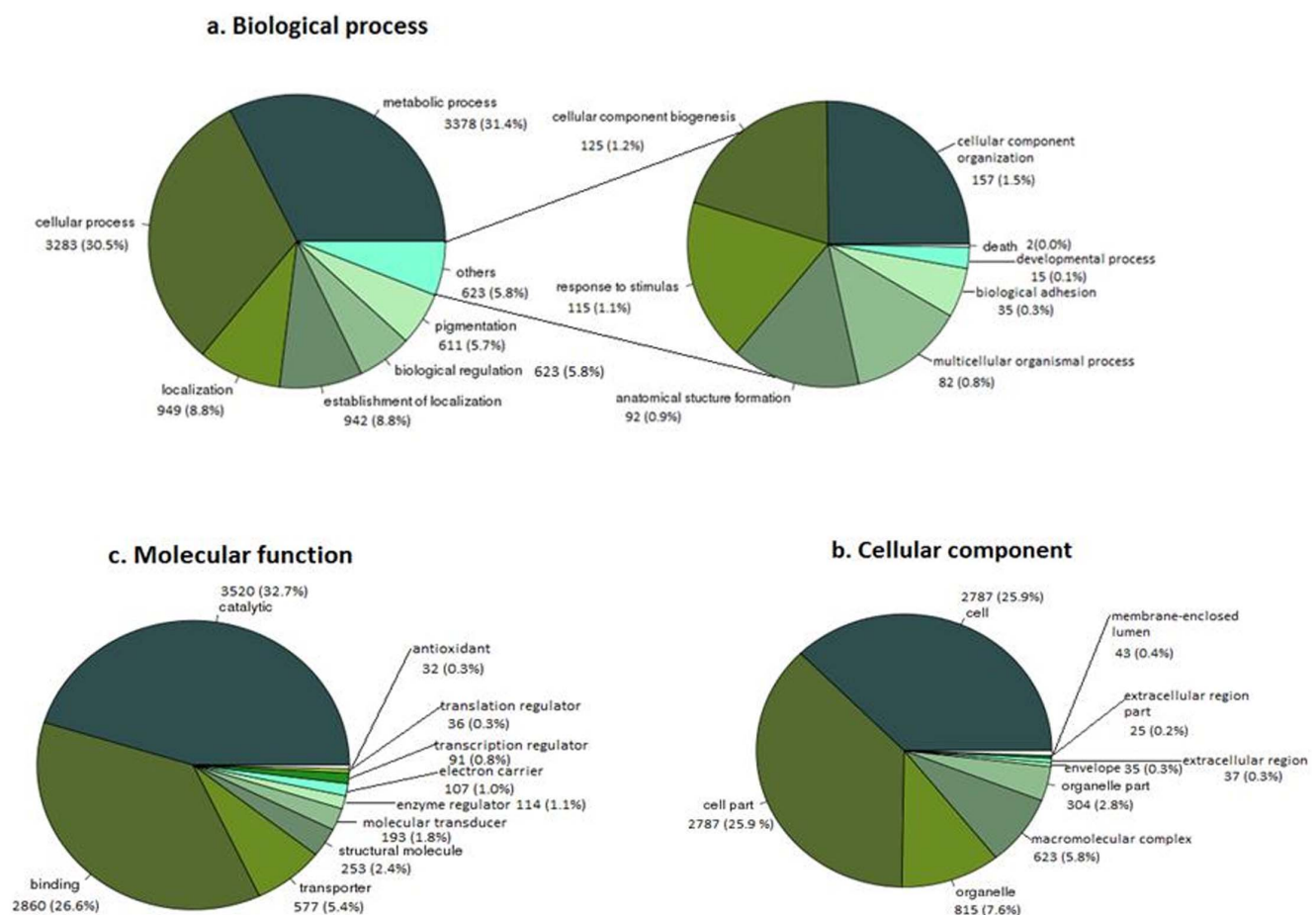


Figure 3. Functional classification of *H. avenae* transcripts using Gene ontology (GO) terms, a. Biological process, b. Molecular function, c. Cellular component.

doi:10.1371/journal.pone.0096311.g003

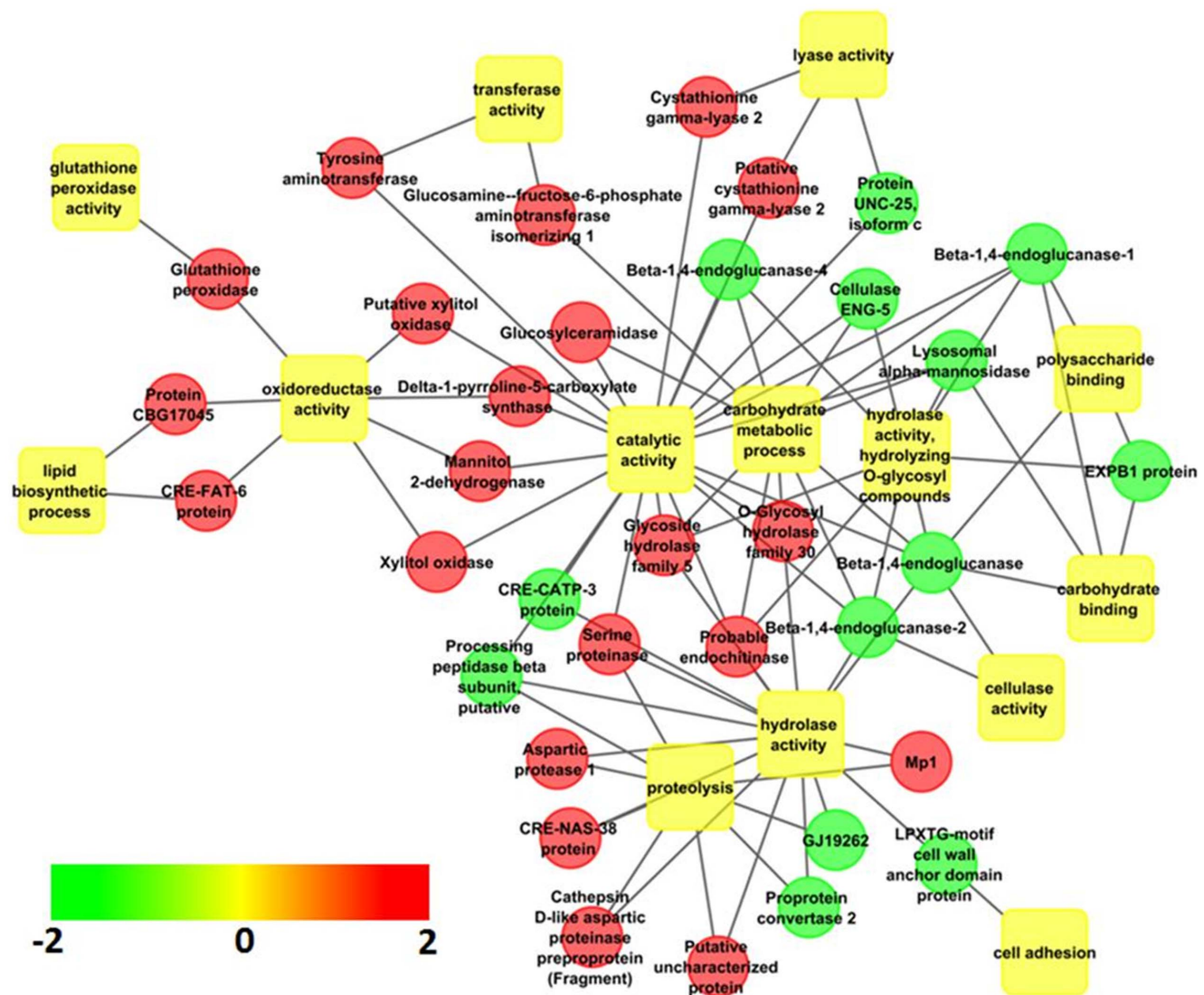


Figure 4. Hierarchical layout of significantly enriched biological processes and key regulatory genes in *H. avenae*.
doi:10.1371/journal.pone.0096311.g004

Peritrophin-A domain containing protein, *CBN-FAT-5 protein*, *dorsal gland cell-specific expression protein*, *CRE*, *CBN-AQP-4*, *CLAVATA*) and the remaining 15 were up-regulated in the J2 stage (*Pectate lyase*, *Serine protease*, β -1-4-endoglucanase, *Annexin A8*, *Putative gland protein G12H04*, *CAEBREN_17602*, *CAEBREN_19301*, *IMG5_133010*, *CLF_100574*, *CBG10845*, *OsL_22746*, *Radixin*, *Dynein light chain 2*, *pfam02414*).

Transcripts for *CBN-AQP-4*, *Spectraplaklin*, *Ran-BPM*, *CRE* and *Superoxide dismutase* showed very high expression in females with the fold change values 22693, 1152, 643, 471 and 230 respectively compared to J2 (Figure 5), whereas *Annexin A8*, *pfam02414*, β -1-4-endoglucanase, *Putative gland protein G12H04*, *OsL_22746* and *Pectate lyase* showed high expression in J2 and the respective fold change values were 1251, 501, 278, 170, 109, 87 compared to female (Figure 6). qRT-PCR results were in agreement with those obtained by digital gene expression.

Secretory peptides, CAZymes and genes involved in parasitism

Analysis to predict secreted proteins in *H. avenae* was undertaken using three approaches. Transcripts with predicted signal peptides

lacking trans-membrane domains were identified by TMHMM and putative secretory sequences similar to plant host species were excluded. This analysis revealed that 1,471 of the contigs had signal peptides but lacked a trans-membrane helix (TMH). Of these, 191 showing significant similarity with the host were excluded from the analysis (Table S7). Among these homologues, 13 belonged to known effectors of *Heterodera* spp (Table 5). Homologues to some potentially interesting effectors were identified including a fatty acid retinoid binding protein RAN-BP-like protein (*Gp-rbp-1*), esophageal gland cell secretory protein (*Hg-hsp-11*), pectate lyase (*Hg-PL-2*), esophageal gland cell secretory protein 1 (*Hg-hsp-1*) and *Annexin* (*Hg-4C10*).

Transcripts with homology to genes that might be involved in the parasitic interaction were identified such as some that modify plant cell walls (CAZymes) [48]. Identification of CAZymes was done using the CAZymes Analysis Toolkit (CAT) that classified 962 transcripts in this category (Table 6, Table S8). The number of transcripts representing different CAZyme families were glycosyl hydrolases - GHs (39), glycosyl transferases - GTs (51), pectate lyases - PLs (2), carbohydrate esterases - CEs (6) and carbohydrate binding modules - CBMs (14) (Table S8). A

Table 4. Statistically significant regulated Gene Ontology categories and pathways in *H. avenae*.

Category	Term	Highly Expressed*	Z Score	p Value	q Value
Cellular component	integral to membrane (GO:0016021)	36	1.87	0.23	1.00
Biological process	proteolysis (GO:0006508)	11	1.84	0.16	1.00
Molecular function	calcium ion binding (GO:0005509)	10	1.68	0.20	1.00
Biological process	G-protein coupled receptor protein signaling pathway (GO:0007186)	7	2.15	0.08	1.00
Molecular function	hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553)	11	7.34	0.00	0.11
Molecular function	motor activity (GO:0003774)	4	2.88	0.04	1.00
Molecular function	carbohydrate binding (GO:0030246)	6	5.63	0.00	0.17
Cellular component	collagen (GO:0005581)	6	7.03	0.00	0.11
Biological process	fatty acid biosynthetic process (GO:0006633)	2	1.95	0.13	1.00
Biological process	response to oxidative stress (GO:0006979)	4	4.68	0.00	0.08
Molecular function	peroxidase activity (GO:0004601)	3	3.55	0.02	0.76
KEGG	ABC transporters:KEGG-ko02010	3	5.52	0.00	0.24
KEGG	Butanoate metabolism:KEGG-ko00650	3	3.77	0.02	1.00

*Based on transcript quantitation and RPKM method.
doi:10.1371/journal.pone.0096311.t004

comparative study of the abundance of CAZymes between *H. avenae* and *M. incognita* suggested that *M. incognita* possessed more genes for hydrolytic activity. In general, CBMs are the non-catalytic domains that appends to glycoside hydrolase enzymes that degrades polysaccharides [49–51] (Table S8). Sixty-four CBM families have been reported in the CAZy database [52] out of which seven (2, 5, 13, 14, 18, 20 and 50) were found in the *H. avenae* transcriptome.

Repeat elements in *H. avenae* transcripts and identification of SSRs

The transcriptome data was also used to identify repeat elements present in *H. avenae* since there is no information on genome-wide repeats for this species. The Repeat Masker program [36] was used to identify different repetitive elements among the transcripts. Approximately 3% of the total transcripts were found to be encoded by different repetitive elements (Table S9). Low complexity regions encoded most transcripts from repetitive elements. A total of 91 retroelements were found amongst transcripts, with 43 long interspersed repeat elements (LINEs), though surprisingly, there were no short interspersed repeat elements (SINEs). Among retroelements, the number of long terminal repeats (LTRs) was slightly higher (48) than non-LTR elements. Also, 28 penelope-like elements distinct from LTR and non-LTR elements were found [53]. In addition, DNA transposons of different classes, simple repeats and small RNAs were identified. For comparison, the *C. elegans* genome contains about 524 SINE elements of which 46% are encoded in the intronic regions [54].

Perl script MISA was used to identify SSRs and generated 1,422 SSRs found in 1,125 transcripts, with a frequency of one SSR per 9.33 kb of the sequence (Table S10). Tri-nucleotide SSRs represented the largest fraction (68.3%) followed by mono-nucleotide (22.6%) and di-nucleotide (6.6%) repeats. Only a few tetra- (33) and penta-nucleotide repeat (1) SSRs were identified in the *H. avenae* transcripts.

Discussion

In this study, we have sequenced and annotated the transcriptome of two stages of *H. avenae* after deep sequencing [55].

The combined assembled contigs from RNA-Seq of J2s and FFs generated 27,765 contigs with N50 of 1,028 bp, for which BLAST searches yielded 37% (10,454) with significant homologies to previously annotated genes in standard databases. A comparison with the EST dataset of NEMABASE4 gave 1,839 clade-specific unique hits, most from the Heteroderidae that increased transcript annotation by a further 6%; with 57% of the contigs uncharacterized. Comparative analyses of contigs of the J2/FF combined transcriptome with sequence data from other classes of nematodes also revealed homologies with free living and animal parasitic nematodes, and the available *H. avenae* contigs shared 38.9% similarity with those of the potato cyst nematode, *G. pallida*. Assignment of GO terms categorised 10,751 transcripts to putative functions.

Our analysis shows that the *H. avenae* transcriptome encodes messages most similar to those thought to be involved in parasitism by *G. pallida*. To be a successful plant pathogen *H. avenae* must locate a host plant root, enter it using its mouth stylet, and migrate intracellularly from cell-to-cell before inducing the feeding site syncytium. As has been found for other plant endoparasitic nematodes, a range of cell wall degrading enzymes have been identified, possibly required to modify plant cell walls during migration, feeding or syncytium formation. Thus, the cell wall modifying CAZymes of *H. avenae* could be expected to be involved in these processes. Thirteen GH families were found in *H. avenae* transcripts: amongst them, GH5 cellulases are most common to all the PPNs except for *B. xylophilous* where GH45 is present. Two β -1,4 endoglucanases of *G. rostochiensis* and *H. glycines* are thought to be involved in J2 migration to the final feeding site [56]. Some endoparasitic nematode ‘parasitism’ genes, particularly those potentially involved in plant cell wall modification, share a high degree of similarity with genes from bacteria and fungi. This suggests that horizontal gene transfer (HGT) has occurred during evolution to parasitism, since such genes are not present in free living nematodes [56]. The results obtained from *H. avenae* are consistent with this view.

Cyst nematodes such as *H. avenae* must form multinucleate syncytial feeding sites by altering the differentiation of a group of cells which fuse together accompanied by local wall degradation. The nuclei in syncytia become endo-polyplod, and related to this we have identified 15 transcripts which may be involved in cell

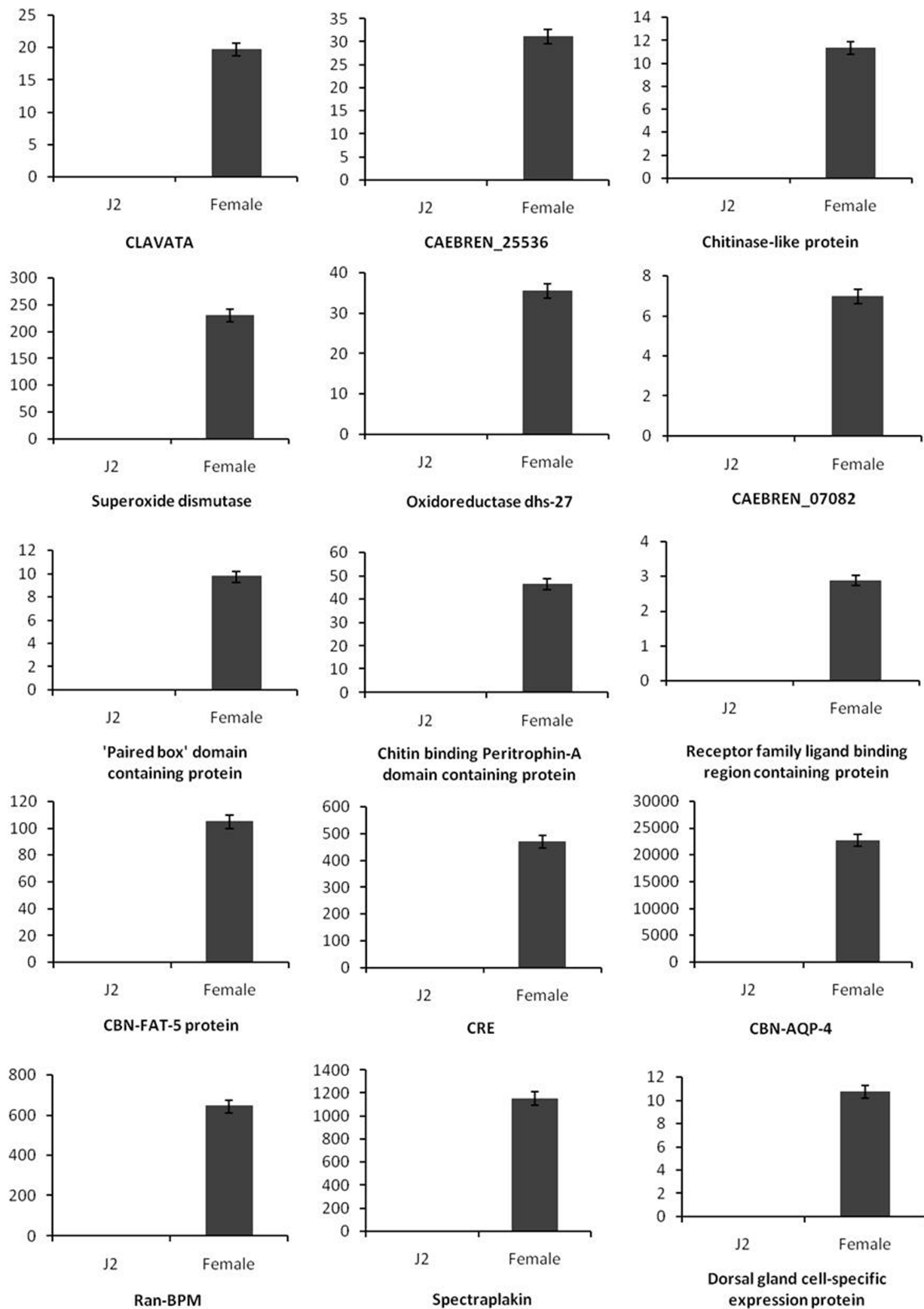


Figure 5. Confirmation of up regulated genes in the female stage of *H. avenae* by qRT-PCR. The Y-axis represents the log2 fold change values. Error bars show \pm SD among the biological replicates. 18S rRNA was used as an internal control gene and fold change was calculated using $2^{-\Delta\Delta CT}$ method.
doi:10.1371/journal.pone.0096311.g005

cycle regulation and development of endo-polyploidy. Similarly, transcripts for *Expansins*, which are cell wall loosening proteins involved in growth and cell wall disassembly and in syncytium expansion [57] are also present in *H. avenae*. Surprisingly, *Chorismate mutase*, which is thought to be involved in evading plant defenses against invading nematodes was not found in our *H. avenae* transcriptome data [1,58–60].

Genes important for the interaction between the nematode and its host are likely to be secreted into host tissues or cells. To predict putative secreted proteins encoded by *H. avenae* we identified transcripts with signal peptides with no trans-membrane helices/domain. Some of the most interesting contigs with these characteristics having no matches in any free living nematodes, could be involved in the interaction of *H. avenae* with its host. Out of the 1,280 transcripts identified to be secretory, 602 transcripts had no matches. Since parasitic nematodes have to suppress different host defense responses for their survival, they may achieve these using different mechanisms. The C-type lectin found in *H. avenae* is an interesting gene that could be involved in overcoming the host defenses as reported in animal parasitic nematodes [61,62], since survival of *H. glycines* in host plants was significantly reduced when RNAi was used to down-regulate a C-type lectin [63]. Another important gene family, SPRYSECs has also been identified in *H. avenae* [64]. Eight types of SPRY domain-containing transcripts were identified, that might help to counter host immune responses [64,65]. Similarly, several animal parasitic nematodes secrete antioxidant enzymes that may provide protection against host-derived reactive oxygen species (ROS). Presence of such antioxidant enzymes in plant parasitic nematodes may similarly provide protection against plant defenses [66]. We have found transcripts for the enzyme *glutathione reductase* that could provide protection to *H. avenae* against ROS. Venom-allergen-like proteins (VAPs or VALs) are highly conserved proteins released by parasitic nematodes suggesting their importance (Table 5) [67]. However, attempts to knock down VAPs in animal parasitic nematodes with RNAi have not been successful, but Lozano (personal communication) used transgenic potato plants to knock down VAPs in *G. rostochiensis*, and this significantly reduced the infectivity of nematodes, suggesting that VAPs are required for successful parasitism.

Modulation of host protein turnover rate may also be important for successful parasitism. Ubiquitination has been implicated in the regulation of many processes in plants, such as innate immunity, cell death, cell cycle regulation, hormone signaling and circadian rhythms. Plant pathogens may hijack the ubiquitination system of the host to modulate cellular processes. Presence of the ubiquitination complex components in pharyngeal glands of PPNs indicates that they may also exploit the host's ubiquitination system [68], and this is consistent with the presence of ubiquitination complex components found in *H. avenae*.

Since *H. avenae* may survive in the soil for several years, it is of interest to understand its nutritional needs and mechanism of survival [69]. Lipids are the primary source of energy in eggs of PPNs, and lipid droplets are prominent in early J2s and FFs. Interestingly, we have identified many transcripts that mapped to the lipid biosynthesis pathway (20). PPNs secrete a specific class of fatty-acid and retinol-binding (*FAR*) proteins that may interfere with lipid-based defenses by inhibiting the production of jasmonic acid [70] in the host, and a *FAR* protein homologue is present in

H. avenae. Similarly, *Annexins* (*Hg4F01*) are phospholipid-binding proteins in nematodes with possible immunomodulatory properties. *Annexins* bind phospholipids, the main components of cell membranes, in a calcium-dependent manner. *Annexins* are expressed in the pharyngeal and the amphidial glands of plant parasitic cyst nematodes [71] (Table 5).

Genome wide RNAi phenotypic studies in *C. elegans* have provided information on genes vital for growth and development. Comparison of *H. avenae* contigs with genes of *C. elegans* lethal RNAi phenotypes gave 4,147 homologous contigs which are expressed in both J2s and FFs of *H. avenae*. These genes are involved in various molecular functions that directly affect movement, pharyngeal pumping, lipid metabolism, embryogenesis, larval development, moulting, growth, reproduction and defence response to chemicals. An example of such phenotype in PPNs is the silencing of two genes involved in movement, *pat-10* and *unc-87*, in J2s of *P. thornei* which reduced reproduction by 77–81% on carrot mini discs. The phenotypic effect after feeding dsRNA to these nematodes was abnormal behaviour including twitching, slow movement, repeated banging of the head against the body and loss of orientation [72].

Genes involved in the RNAi pathway are well characterised in many nematode species, although there is less knowledge of the pathways involved in RNAi in PPNs compared to *C. elegans*. A total of 30 genes involved in the RNAi pathway were found amongst the *H. avenae* transcripts, including a putative nuclear AGO (Argonaute) *NRDE-3* not previously identified in a PPN (Table 3) [44] and this needs to be further investigated.

Nematodes have a primitive central nervous system (CNS) and use signaling processes involving neuropeptides [45] that have a diverse role in the function and development of nervous system [73]. They not only act as neuromodulators, but also as primary transmitters in invertebrate nervous systems [73,74]. To date 109 neuropeptide genes have been identified in *C. elegans* [73,75,76]. Based on their conserved motifs, these are divided into three classes: the FMRFamide-like peptide (FLP) gene family, INS genes that encode insulin-like peptides and peptides derived from neuropeptide-like protein (NLP) genes which have no sequence similarity to the other two classes [77]. In the *H. avenae* transcriptome data, we found several neuropeptide genes belonging to all the three classes. These included *flp-2*, 3, 11 and 18, four *flp*-receptors, nine NLPs and several INS genes such as Insulin-like growth factor 2. The function of *flp* gene products is better understood in *C. elegans* than in PPNs [44]. Expression studies in *C. elegans* using *flp-18* reporter gene constructs revealed its expression in the specific interneurons AVA, AIY and RIG, the motor neuron RIM and pharyngeal neurons M2 and M3 [78]. *flp-18* mutants in *C. elegans* were defective in chemosensation, foraging, dauer formation and fat accumulation and also exhibited decreased oxygen consumption. *flp-18* mutants in *C. elegans* were defective in chemosensation, foraging, dauer formation and fat accumulation and also exhibited decreased oxygen consumption. Host delivered RNAi of *flp-18* significantly reduced the infection and multiplication of *M. incognita* [79]. Similar effects could be envisaged from knock down of these genes in *H. avenae*.

In addition, to confirm the *in silico* quantitation, a selection of 30 of the putatively differentially expressed genes were checked by qRT-PCR and their relative levels of expression validated. High expression of pectate lyase and β -1, 4 *endoglucanase*, *Annexin* and

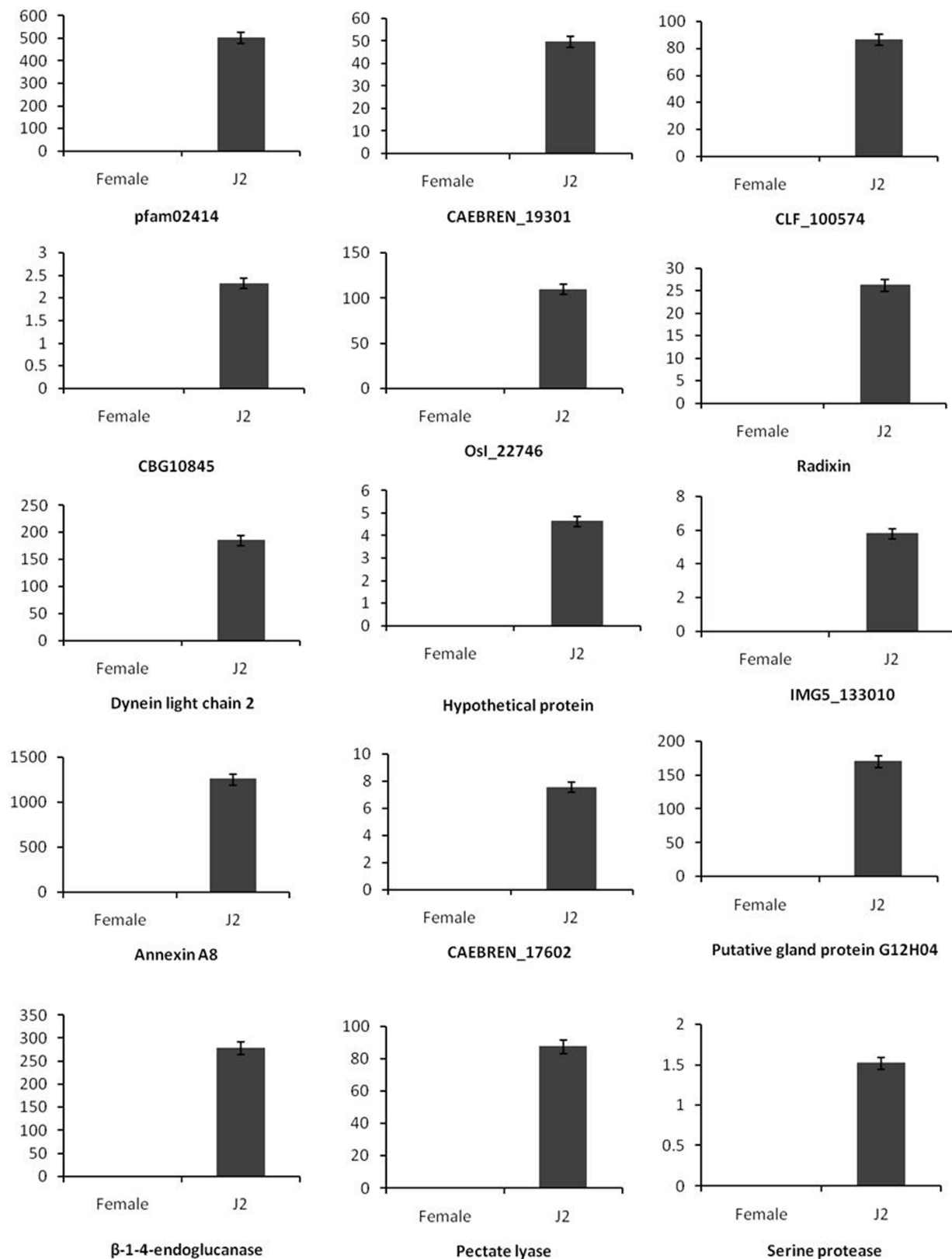


Figure 6. Confirmation of up regulated genes in juvenile stage of *H. avenae* by qRT-PCR. The Y-axis represents the log2 fold change values. Error bars show \pm SD among the biological replicates. *18S rRNA* was used as an internal control gene and fold change was calculated by using $2^{-\Delta\Delta CT}$ method.

doi:10.1371/journal.pone.0096311.g006

Table 5. Important genes identified in *H. avenae* based on best BLASTX hits.

Accession No.	Descriptions	Species	e-Value	%Identity	Reference
ABV54446	GHF5 - endo-1,4-beta-glucanase precursor	<i>Radopholus similis</i>	7.00E-23	46	[83,84]
ADY41344	B 1,4-alpha-glucan-branching enzyme	<i>Ascaris suum</i>	8E-133	77	[83]
ADL29728	Expansin	<i>Heterodera glycines</i>	7.00E-20	75	[57]
ADD82848	Pectate lyase	<i>Heterodera avenae</i>	2.00E-133	97	[85]
ACO55952	beta-1,4-endoglucanase	<i>Heterodera avenae</i>	4E-10	81	[84]
ACV31368	expansin B2	<i>Globodera rostochiensis</i>	0.003	90	[57]
ACN93668	Annexin 4F01	<i>Heterodera schachtii</i>	5.00E-07	58	[86]
Q9NFS0	Putative hypodermis secreted protein	<i>Globodera rostochiensis</i>	3E-57	71	[87]
AAK60209	venom allergen-like protein-1	<i>Heterodera glycines</i>	2.00E-09	83	[88]
NP_001123113	Cysteine Protease related family member (cpr-6)	<i>Caenorhabditis elegans</i>	2.00E-08	77	[89]
CAD38523	Glutathione peroxidase	<i>Heterodera glycines</i>	0.0000004	65	[90]
ACO35733	Ran-Binding protein (RBP-1)	<i>Globodera pallida</i>	7E-11	43	[91]
A4Q9L4	Transthyretin-like protein 4	<i>Radopholus similis</i>	6.00E-69	79	[92]
Q9BJ52	Putative zinc finger protein	<i>Heterodera glycines</i>	6E-38	81	Unpublished
Q86DU4	Ubiquitin extension	<i>Heterodera schachtii</i>	3.00E-14	59	[93]
P53021	(Major sperm protein-1) MSP-1	<i>Globodera rostochiensis</i>	2.00E-42	83	Unpublished
I3VB56	Calreticulin	<i>Radopholus similis</i>	1.00E-76	91	[94]
Q93142	FAR-1	<i>Brugia malayi</i>	9e-17	35	[95]
I6ZQH6	SPRYSEC-8	<i>Globodera rostochiensis</i>			[65]
AF159590	Hsp-11	<i>Heterodera glycines</i>	1E-29	54	[96]
ADY48649	Ubiquitin-conjugating enzyme E2	<i>Ascaris suum</i>	4.50E-81	43	[97]
ADY39814	E3 ubiquitin-protein ligase	<i>Ascaris suum</i>	8.70E-102	55	[97]
ADY41696	Ubiquitin carboxyl-terminal hydrolase 8	<i>Ascaris suum</i>	9.41E-113	51	[97]
Q7YWJ0	MSP-21	<i>Meloidogyne incognita</i>	3.00E-124	46	[98]

doi:10.1371/journal.pone.0096311.t005

putative gland protein *G12H04* supported the suggestion that in the infective J2s some of these gene products may be required for host invasion, whereas genes highly expressed in FFs are more

likely to be involved in other metabolic activities such as reproduction: these results are consistent with similar earlier observations for cyst nematodes [80]. Up-regulation of *CLAVATA*

Table 6. *Heterodera avenae* enzymes with predicted cell wall-degrading activities, compared with those in other nematodes.

Substrate	Cellulose			Xylan	Pectin	Arabinan	
Family	GH5	GH45	GH27	GH30	GH28	PL3	GH43
<i>H. avenae</i>	16	0	5	0	0	2	2
<i>M. incognita</i>	21	0	4	6	2	30	2
<i>B. xylophilus</i>	0	11	0	0	0	15	0
<i>C. elegans</i>	0	0	0	0	0	0	0
Substrate	1,3-glucan		Chitin				
Family	GH16		GH18		GH19		GH20
<i>H. avenae</i>	0		0		1		1
<i>M. incognita</i>	0		3		2		2
<i>B. xylophilus</i>	6		9		2		7
<i>C. elegans</i>	0		38		6		5

GH- Glycosidehydrolases,GT-Glycosyl transferases,CE-Carbohydrate esterases, PL - Polysaccharide lyases.

doi:10.1371/journal.pone.0096311.t006

in FFs could indicate their longer involvement with maintaining feeding site integration [81].

About 3% of the total transcripts encoded different repetitive elements. This may be due to lack of expression of repetitive elements or difficulty in assembling short reads containing repeat elements. However, many SSRs were identified in *H. avenae*, with more tri-nucleotide SSRs than di-nucleotide SSRs, and this agrees with observations obtained from genomic sequencing of *M. incognita* [82]. The identification of SSRs may be useful for the development of large sets of markers which would facilitate linkage mapping studies and population genetics research on *H. avenae*.

In conclusion, we have undertaken the first *de novo* analysis of the transcriptomes of two life stages of *H. avenae*, and identified some genes that may be important in plant parasitism. A comparative analysis of gene expression between the two stages of the parasite provides additional information on some genes that are likely to be involved either in parasitism or nematode metabolism. These data on the *H. avenae* transcriptome should be a valuable resource for future genomic studies on these economically important plant parasitic nematodes.

Supporting Information

Figure S1 Quality check of RNA by Bioanalyzer (Agilent), RNA area: 150.0, RNA concentration: 112 ng/μl, rRNA Ratio [28 s/18 s]:1.4, RNA integrity number (RIN):8.
(TIF)

Figure S2 Sequence length distribution of *H. avenae* assembled contigs.
(TIF)

Table S1 Primer information and BLASTN search results for the genes used for quantitative real time PCR.
(XLSX)

Table S2 Complete BLAST hit table against nr GENBANK, Refseq, Swissprot, EMBL, DDBJ, PIR and RCSB databases and transcript quantitation.
(XLSX)

References

- Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, et al. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature biotechnology* 26: 909–915.
- Yan G, Smiley RW (2010) Distinguishing *Heterodera filipjevi* and *H. avenae* using polymerase chain reaction-restriction fragment length polymorphism and cyst morphology. *Phytopathology* 100: 216–224.
- Pankaj DSCa (1998) Nematodes affecting wheat and rice and their management.. In: *Pathological problems of economic crop plants and their management*. Publ. Scientific Publishers, Jodhpur, India. 95–117 p.
- Brown RH, Kerry BR (1987) *Principles and practice of nematode control in crops*. Sydney N.S.W.; Orlando Fla.: Academic Press. xiv, 447 p. p.
- Mokabli A, Valette S, Gauthier JP, Rivoal R (2001) Influence of temperature on the hatch of *Heterodera avenae* Woll. populations from Algeria. *Nematology* 3: 171–178.
- Gheysen G, Fenoll C (2002) Gene expression in nematode feeding sites. *Annual review of phytopathology* 40: 191–219.
- JONES MGK (1981) Host cell responses to endoparasitic nematode attack: structure and function of giant cells and syncytia. *Annals of Applied Biology* 97: 353–372.
- Wyss U, Grundler FMW (1992) Feeding-Behavior of Sedentary Plant Parasitic Nematodes. *Netherlands Journal of Plant Pathology* 98: 165–173.
- (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282: 2012–2018.
- Ghedini E, Wang S, Spiro D, Caler E, Zhao Q, et al. (2007) Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317: 1756–1760.
- Elsworth B, Wasmuth J, Blaxter M (2011) NEMBASE4: the nematode transcriptome resource. *International journal for parasitology* 41: 881–894.
- Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, et al. (2008) Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc Natl Acad Sci U S A* 105: 14802–14807.
- Alkharouf NW, Klink VP, Matthews BF (2007) Identification of *Heterodera glycines* (soybean cyst nematode [SCN]) cDNA sequences with high identity to those of *Caenorhabditis elegans* having lethal mutant or RNAi phenotypes. *Exp Parasitol* 115: 247–258.
- Vanholme B, Mitreva M, Van Crielinge W, Logghe M, Bird D, et al. (2006) Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*. *Parasitol Res* 98: 414–424.
- Cotton JA, Lilley CJ, Jones LM, Kikuchi T, Reid AJ, et al. (2014) The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol* 15: R43.
- Haegeman A, Joseph S, Gheysen G (2011) Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology. *Mol Biochem Parasitol* 178: 7–14.
- Nicol P, Gill R, Fosu-Nyarko J, Jones MG (2012) *de novo* analysis and functional classification of the transcriptome of the root lesion nematode, *Pratylenchus thornei*, after 454 GS FLX sequencing. *Int J Parasitol* 42: 225–237.
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57–63.
- Rao U, Sharma A, Tyagi N, Banakar P, Kumar M (2013) Characterization of genetic homogeneity of an Indian population of cereal cyst nematode, *Heterodera avenae* using sequencing and PCR-RFLP of ribosomal DNA. *Bioinformation* 9: 67–71.
- Miller HC, Biggs PJ, Voelckel C, Nelson NJ (2012) *De novo* sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (*Sphenodon punctatus*). *BMC Genomics* 13: 439.

Table S3 (A) BLASTN result against NEMABASE4 data (B) 1,839 hits did not show any homologues in NR database.
(XLSX)

Table S4 Complete list of species used for nematode specific BLASTN against NEMABASE4 data.
(XLSX)

Table S5 RNAi lethal phenotype homologues found in the *H. avenae* transcriptome.
(XLSX)

Table S6 (A) Summary of KEGG pathway information for *H. avenae* transcriptome (B) KEGG pathways for 3,601 transcripts of *H. avenae* transcriptome.
(XLSX)

Table S7 Transcripts potentially secreted by *H. avenae*.
(XLSX)

Table S8 Details of all the CAZymes in the *H. avenae* transcriptome with blast and enzymes commission result.
(XLS)

Table S9 Expressed repeat elements identified in *H. avenae* transcriptome.
(XLSX)

Table S10 Details of the expressed Simple Sequence Repeats identified in *H. avenae* transcripts.
(XLSX)

Acknowledgments

We would like to acknowledge Prof. Alok Bhattacharya for his guidance. We are also thankful to Dr. Satyapal Bishnoi and fellow lab members for biological material collection. M.G.K. Jones and Dr J. Fosu-Nyarko for helpful comments.

Author Contributions

Conceived and designed the experiments: UR. Performed the experiments: MK NPG TR PK PB RNS. Analyzed the data: MK NPG TR PK RNS. Wrote the paper: UR MGKJ MK PK NPG.

21. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829.
22. Langmead B (2010) Aligning short sequencing reads with Bowtie. *Current protocols in bioinformatics/editorial board, Andreas D Baxevanis, [et al] Chapter 11: Unit 11 17.*
23. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
24. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009.
25. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2005) GenBank. *Nucleic Acids Res* 33: D34–38.
26. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic acids research* 28: 45–48.
27. Kulikova T, Akhtar N, Aldebert P, Althorpe N, Andersson M, et al. (2007) EMBL Nucleotide Sequence Database in 2006. *Nucleic acids research* 35: D16–20.
28. Tateno Y, Imanishi T, Miyazaki S, Fukami-Kobayashi K, Saitou N, et al. (2002) DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic acids research* 30: 27–30.
29. Barker WC, Garavelli JS, Huang H, McGarvey PB, Orcutt BC, et al. (2000) The protein information resource (PIR). *Nucleic acids research* 28: 41–44.
30. Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, et al. (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic acids research* 39: D392–401.
31. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
32. Park BH, Karpinetz TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20: 1574–1584.
33. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785–786.
34. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567–580.
35. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S (2004) Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* 17: 349–356.
36. Smit A, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0.
37. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 27: 573–580.
38. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359.
39. Zambon AC, Gaj S, Ho I, Hanspers K, Vranizan K, et al. (2012) GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* 28: 2209–2210.
40. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25: 402–408.
41. Mitreva M, Wendl M, Martin J, Wylie T, Yin Y, et al. (2006) Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol* 7: R75.
42. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, et al. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806–811.
43. Deng W, Nickle DC, Learn GH, Maust B, Mullins JI (2007) ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics* 23: 2334–2336.
44. Dalzell JJ, McVeigh P, Warnock ND, Mitreva M, Bird DM, et al. (2011) RNAi Effector Diversity in Nematodes. *Plos Neglected Tropical Diseases* 5.
45. Kimber MJ, Fleming CC (2005) Neuromuscular function in plant parasitic nematodes: a target for novel control strategies? *Parasitology* 131 Suppl: S129–142.
46. Davis EL, Hussey RS, Mitchum MG, Baum TJ (2008) Parasitism proteins in nematode-plant interactions. *Current opinion in plant biology* 11: 360–366.
47. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research* 38: D355–360.
48. Bakhietia M, Urwin PE, Atkinson HJ (2007) QPCR analysis and RNAi define pharyngeal gland cell-expressed genes of *Heterodera glycines* required for initial interactions with the host. *Molecular plant-microbe interactions: MPMI* 20: 306–312.
49. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382: 769–781.
50. Reinikainen T, Ruohonen L, Nevanen T, Laaksonen L, Kraulis P, et al. (1992) Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. *Proteins* 14: 475–482.
51. Arantes V, Saddler JN (2010) Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol Biofuels* 3: 4.
52. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37: D233–238.
53. Evgen'ev MB, Arkhipova IR (2005) Penelope-like elements—a new class of retroelements: distribution, function and possible evolutionary significance. *Cytogenet Genome Res* 110: 510–521.
54. Sela N, Kim E, Ast G (2010) The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome biology* 11: R59.
55. Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 12: R112.
56. Popeijus H, Overmars H, Jones J, Blok V, Govers A, et al. (2000) Degradation of plant cell walls by a nematode. *Nature* 406: 36–37.
57. Wiczeorek K, Grundler FM (2006) Expanding nematode-induced syncytia: the role of expansins. *Plant Signal Behav* 1: 223–224.
58. Lambert KN, Allen KD, Sussex IM (1999) Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. *Mol Plant Microbe Interact* 12: 328–336.
59. Jones JT, Furlanetto C, Bakker E, Banks B, Blok V, et al. (2003) Characterization of a chorismate mutase from the potato cyst nematode *Globodera pallida*. *Mol Plant Pathol* 4: 43–50.
60. Bekal S, Niblack TL, Lambert KN (2003) A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Mol Plant Microbe Interact* 16: 439–446.
61. de Boer JM, McDermott JP, Wang X, Maier T, Qui F, et al. (2002) The use of DNA microarrays for the developmental expression analysis of cDNAs from the esophageal gland cell region of *Heterodera glycines*. *Molecular plant pathology* 3: 261–270.
62. Maizels RM, Tetteh KK, Loukas A (2000) *Toxocara canis*: genes expressed by the arrested infective larval stage of a parasitic nematode. *Int J Parasitol* 30: 495–508.
63. Urwin PE, Lilley CJ, Atkinson HJ (2002) Ingestion of double-stranded RNA by parasitic juvenile cyst nematodes leads to RNA interference. *Molecular plant-microbe interactions: MPMI* 15: 747–752.
64. Rehman S, Postma W, Tytgat T, Prins P, Qin L, et al. (2009) A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR protein from a susceptible tomato. *Molecular plant-microbe interactions: MPMI* 22: 330–340.
65. Postma WJ, Slootweg EJ, Rehman S, Finkers-Tomczak A, Tytgat TO, et al. (2012) The effector SPRYSEC-19 of *Globodera rostochiensis* suppresses CC-NB-LRR-mediated disease resistance in plants. *Plant physiology* 160: 944–954.
66. Jones D (2004) Reactive oxygen species and rosacea. *Cutis: cutaneous medicine for the practitioner* 74: 17–20, 32–14.
67. Farias LP, Rodrigues D, Cunha V, Rofatto HK, Faquim-Mauro EL, et al. (2012) *Schistosoma mansoni* venom allergen like proteins present differential allergic responses in a murine model of airway inflammation. *PLoS neglected tropical diseases* 6: e1510.
68. Davis EL, Hussey RS, Baum TJ (2004) Getting to the roots of parasitism by nematodes. *Trends in parasitology* 20: 134–141.
69. Cooper AF, Van Gundy SD (1970) Metabolism of Glycogen and Neutral Lipids by *Aphelenchus avenae* and *Caenorhabditis* sp. in Aerobic, Microaerobic and Anaerobic Environments. *Journal of nematology* 2: 305–315.
70. Prior A, Jones JT, Blok VC, Beauchamp J, McDermott L, et al. (2001) A surface-associated retinol- and fatty acid-binding protein (Gp-FAR-1) from the potato cyst nematode *Globodera pallida*: lipid binding activities, structural analysis and expression pattern. *Biochem J* 356: 387–394.
71. Gao B, Allen R, Maier T, Davis EL, Baum TJ, et al. (2003) The parasitome of the phytonematode *Heterodera glycines*. *Molecular plant-microbe interactions: MPMI* 16: 720–726.
72. Tan JA, Jones MG, Fosu-Nyarko J (2013) Gene silencing in root lesion nematodes (*Pratylenchus* spp.) significantly reduces reproduction in a plant host. *Exp Parasitol* 133: 166–178.
73. Li C, Nelson LS, Kim K, Nathoo A, Hart AC (1999) Neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Annals of the New York Academy of Sciences* 897: 239–252.
74. Li C (2005) The ever-expanding neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Parasitology* 131 Suppl: S109–127.
75. McVeigh P, Leech S, Mair GR, Marks NJ, Geary TG, et al. (2005) Analysis of FMRFamide-like peptide (FLP) diversity in phylum Nematoda. *International journal for parasitology* 35: 1043–1060.
76. Nathoo AN, Moeller RA, Westlund BA, Hart AC (2001) Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. *Proceedings of the National Academy of Sciences of the United States of America* 98: 14000–14005.
77. Huxson SJ, Mertens I, Janssen T, Lindemans M, Schoofs L (2007) Neuropeptidergic signaling in the nematode *Caenorhabditis elegans*. *Progress in neurobiology* 82: 33–55.
78. Rogers C, Reale V, Kim K, Chatwin H, Li C, et al. (2003) Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1. *Nature neuroscience* 6: 1178–1185.
79. Papolu PK, Gantasala NP, Kamaraju D, Banakar P, Sreevathsa R, et al. (2013) Utility of Host Delivered RNAi of Two FMRF Amide Like Peptides, flp-14 and flp-18, for the Management of Root Knot Nematode, *Meloidogyne incognita*. *PLoS One* 8: e80603.

80. Bakhetia M, Urwin PE, Atkinson HJ (2007) QPCR analysis and RNAi define pharyngeal gland cell-expressed genes of *Heterodera glycines* required for initial interactions with the host. *Mol Plant Microbe Interact* 20: 306–312.
81. Replogle A, Wang J, Paolillo V, Smeda J, Kinoshita A, et al. (2013) Synergistic interaction of CLAVATA1, CLAVATA2, and RECEPTOR-LIKE PROTEIN KINASE 2 in cyst nematode parasitism of *Arabidopsis*. *Mol Plant Microbe Interact* 26: 87–96.
82. Castagnone-Sereno P, Danchin EG, Deleury E, Guillemaud T, Malausa T, et al. (2010) Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species *Meloidogyne incognita*. *BMC genomics* 11: 598.
83. Goellner M, Wang X, Davis EL (2001) Endo-beta-1,4-glucanase expression in compatible plant-nematode interactions. *Plant Cell* 13: 2241–2255.
84. Haegeman A, Jones JT, Danchin EG (2011) Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol Plant Microbe Interact* 24: 879–887.
85. Kikuchi T, Shibuya H, Aikawa T, Jones JT (2006) Cloning and characterization of pectate lyases expressed in the esophageal gland of the pine wood nematode *Bursaphelenchus xylophilus*. *Mol Plant Microbe Interact* 19: 280–287.
86. Gao B, Allen R, Maier T, Davis EL, Baum TJ, et al. (2003) The parasitome of the phytoneematode *Heterodera glycines*. *Mol Plant Microbe Interact* 16: 720–726.
87. Qin L, Overmars H, Helder J, Popeijus H, van der Voort JR, et al. (2000) An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. *Molecular plant-microbe interactions: MPMI* 13: 830–836.
88. Gao B, Allen R, Maier T, Davis EL, Baum TJ, et al. (2001) Molecular characterisation and expression of two venom allergen-like protein genes in *Heterodera glycines*. *Int J Parasitol* 31: 1617–1625.
89. Shingles J, Lilley CJ, Atkinson HJ, Urwin PE (2007) *Meloidogyne incognita*: molecular and biochemical characterisation of a cathepsin L cysteine proteinase and the effect on parasitism following RNAi. *Exp Parasitol* 115: 114–120.
90. Jones JT, Reavy B, Smant G, Prior AE (2004) Glutathione peroxidases of the potato cyst nematode *Globodera rostochiensis*. *Gene* 324: 47–54.
91. Vanholme B, De Meutter J, Tytgat T, Van Montagu M, Coomans A, et al. (2004) Secretions of plant-parasitic nematodes: a molecular update. *Gene* 332: 13–27.
92. Matthews BF, Pilitt KL, Klink V (2004) Molecular Characterization of a Soybean Cyst Nematode (*Heterodera glycines*) Homolog of unc-87. *J Nematol* 36: 457–465.
93. Tytgat T, Vanholme B, De Meutter J, Claeys M, Couvreur M, et al. (2004) A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. *Molecular plant-microbe interactions: MPMI* 17: 846–852.
94. Jaubert S, Milac AL, Petrescu AJ, de Almeida-Engler J, Abad P, et al. (2005) In planta secretion of a calreticulin by migratory and sedentary stages of root-knot nematode. *Molecular plant-microbe interactions: MPMI* 18: 1277–1284.
95. Garofalo A, Klager SL, Rowlinson MC, Nirmalan N, Klion A, et al. (2002) The FAR proteins of filarial nematodes: secretion, glycosylation and lipid binding characteristics. *Molecular and biochemical parasitology* 122: 161–170.
96. Wang X, Allen R, Ding X, Goellner M, Maier T, et al. (2001) Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Molecular plant-microbe interactions: MPMI* 14: 536–544.
97. Wang J, Czech B, Crunk A, Wallace A, Mitreva M, et al. (2011) Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. *Genome research* 21: 1462–1477.
98. Huang G, Gao B, Maier T, Allen R, Davis EL, et al. (2003) A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*. *Molecular plant-microbe interactions: MPMI* 16: 376–381.